SPARSE REPRESENTATIONS FOR IMAGE CLASSIFICATION USING QUANTUM D-WAVE 2X MACHINE

Nga Nguyen[1], Amy Larson[1], Carleton Coffrin[2], *and* Garrett Kenyon[1,3]
[1]CCS-3, [2]A-1, *Los Alamos National Laboratory* & [3]*New Mexico Consortium*

D-Wave Debrief, LANL, April 27, 2017

OUTLINE

A. **SPARSE CODING REPRESENTATIONS**
B. **IMPLEMENTATION ON D-WAVE MACHINE**
C. **SPARSE CODING FOR OBJECT DETECTION**
D. **SUMMARY AND FUTURE WORK**

OUTLINE

A. SPARSE CODING REPRESENTATIONS
B. IMPLEMENTATION ON A D-WAVE MACHINE
C. SPARSE CODING FOR OBJECT DETECTION
D. SUMMARY AND FUTURE WORK

A. SPARSE CODING REPRESENTATIONS

Los Alamos NATIONAL LABORATORY EST.1943

Solving a sparse-coding (SC) problem

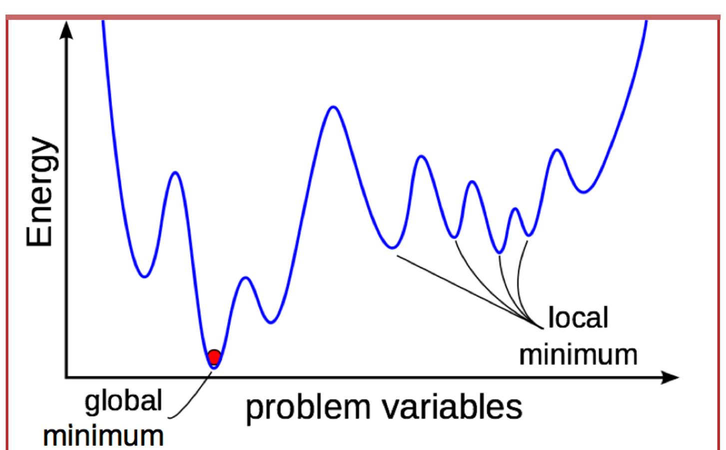
Objective function is of the form:

$$E = \min_{\{\vec{a}, \phi\}} \left[\frac{1}{2} |\vec{I} - \phi \vec{a}|^2 + \lambda ||\vec{a}||_p \right].$$

reconstruction error **Lp-sparseness penalty**

Olshausen and Field, Nature 381, 607 (1996)
Rozell, Johnson, Baraniuk, and Olshausen, Neur. Comp. 20, 2526 (2008)

p=0, the problem is called **L0-norm**



courtesy of D-Wave

- **non-convex problem**
- **NP-hard class**

A. SPARSE CODING REPRESENTATIONS

Los Alamos NATIONAL LABORATORY EST. 1943

- **Solving a sparse-coding (SC) problem**

Objective function is of the form:

$$E = \min_{\{\vec{a}, \phi\}} \left[\frac{1}{2} |\vec{I} - \phi \vec{a}|^2 + \lambda ||\vec{a}||_p \right].$$

reconstruction error

***Lp*-sparseness penalty**

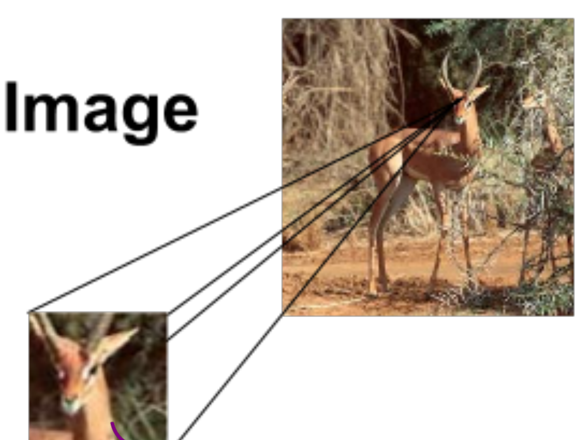
Olshausen and Field, Nature 381, 607 (1996)
Rozell, Johnson, Baraniuk, and Olshausen, Neur. Comp. 20, 2526 (2008)

an example of SC reconstruction

Features (Receptive field)

Activity

Image

() * $(a_1, a_2, \ldots a_n)^T$ = 

ϕ * $\vec{a}$ = $\vec{\hat{I}}$

courtesy of Xinhua Zhang

OUTLINE

A. SPARSE CODING REPRESENTATIONS
B. IMPLEMENTATION ON A D-WAVE MACHINE
C. SPARSE CODING FOR OBJECT DETECTION
D. SUMMARY AND FUTURE WORK

SC ON A QUANTUM D-WAVE MACHINE

- **mapping the sparse-coding problem onto a Quantum Unconstrained Binary Optimization (QUBO):**

D-Wave Hamiltonian:

$$H(h,Q,a) = \sum_i h_i a_i + \sum_{<i,j>} Q_{ij} a_i a_j$$

where $a_i = \{0,1\} \forall i.$

SC ON A QUANTUM D-WAVE MACHINE

- **mapping the sparse-coding problem onto a Quantum Unconstrained Binary Optimization (QUBO):**

D-Wave Hamiltonian:

$$H(h, Q, a) = \sum_i h_i a_i + \sum_{<i,j>} Q_{ij} a_i a_j$$

where $a_i = \{0, 1\} \forall i.$

This mapping is achieved by the relations:

$$h = -\phi^T \overrightarrow{I} + (\frac{1}{2} + \lambda),$$

$$Q = \frac{1}{2} \phi^T \phi.$$

analogous to L0-sparseness penalty [Nguyen and Kenyon, PMES-16 (2016)]

OUTLINE

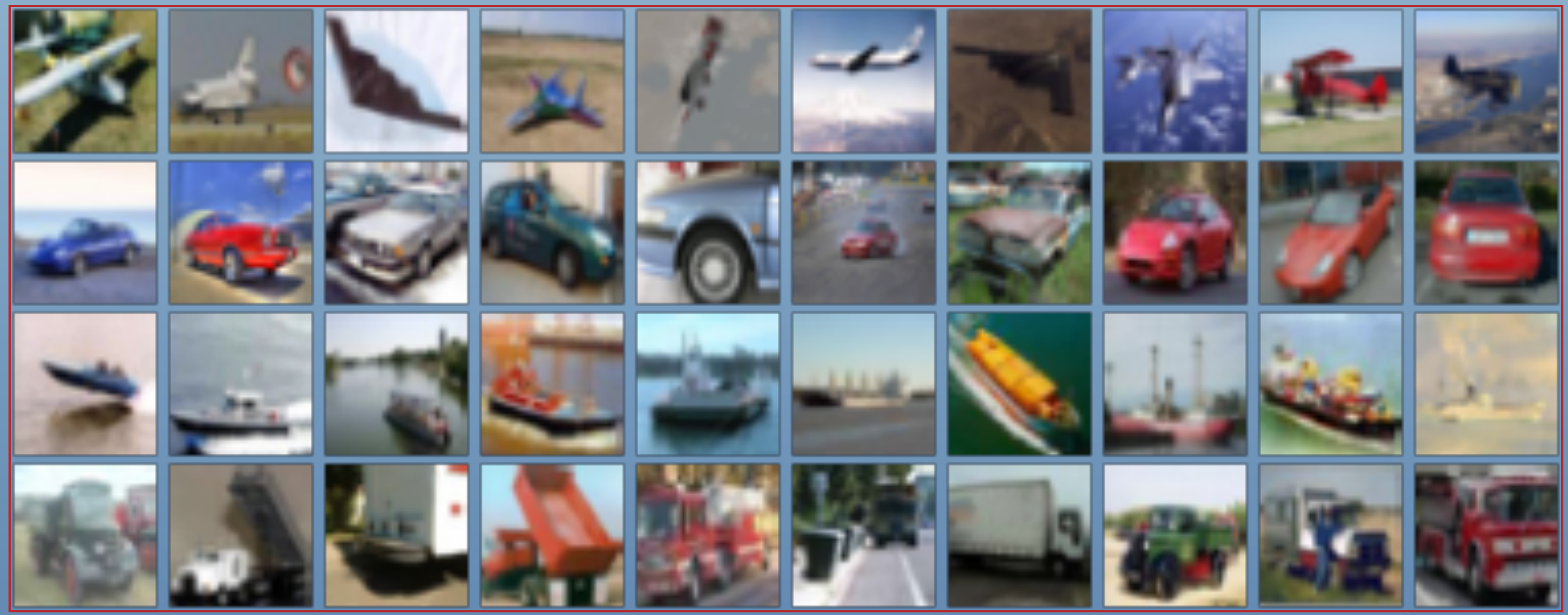
A. SPARSE CODING ON A QUANTUM D-WAVE
B. IMPLEMENTATION ON D-WAVE MACHINE
C. SPARSE CODING FOR OBJECT DETECTION
D. SUMMARY AND FUTURE WORK

DATASET

32x32

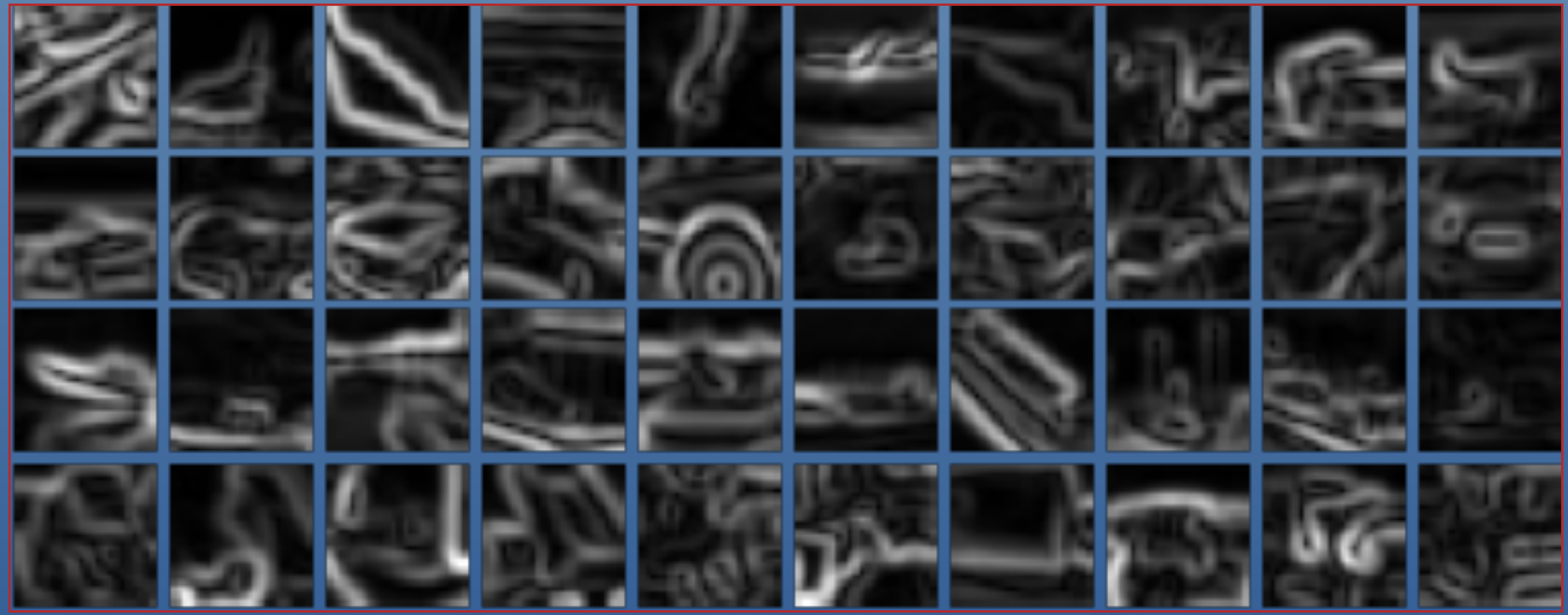
CIFAR-10

airplane
automobile
ship
truck



24x24

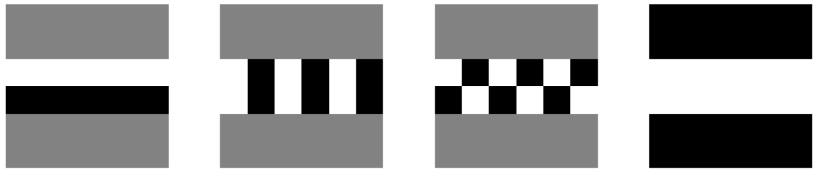
edge detection



Features

8 hand-designed features

"row"

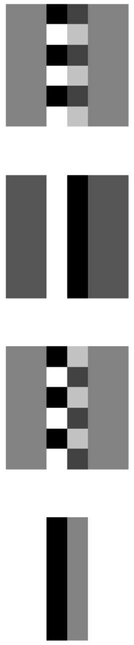


$\{\psi_i\}$

coupling $=<\psi_i,\psi_j>$

orthogonality!

number of features $N_f = 8$

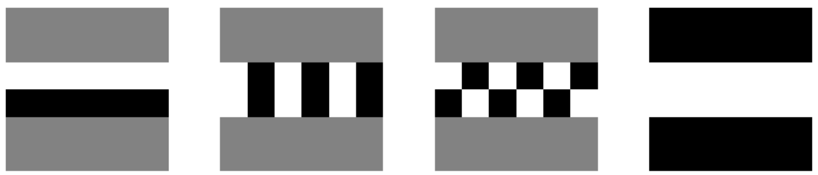


"column"

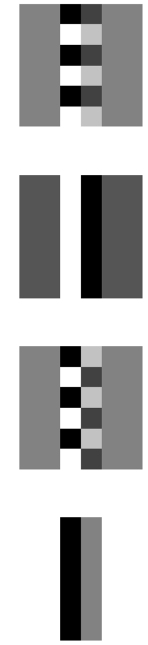
Features

8 hand-designed features

"row"



"column"

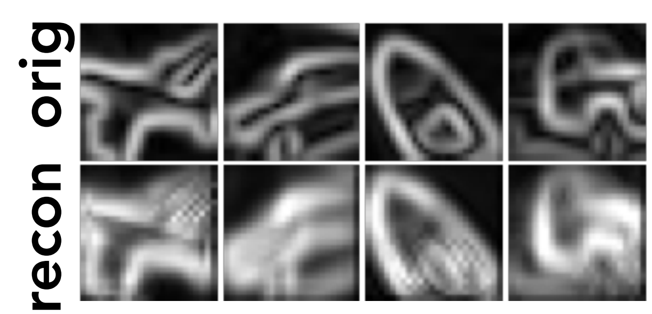


$\{\psi_i\}$

coupling $=< \psi_i, \psi_j >$

orthogonality!

number of features $N_f = 8$



Features

<u>Desire:</u> Randomly generated N_f :

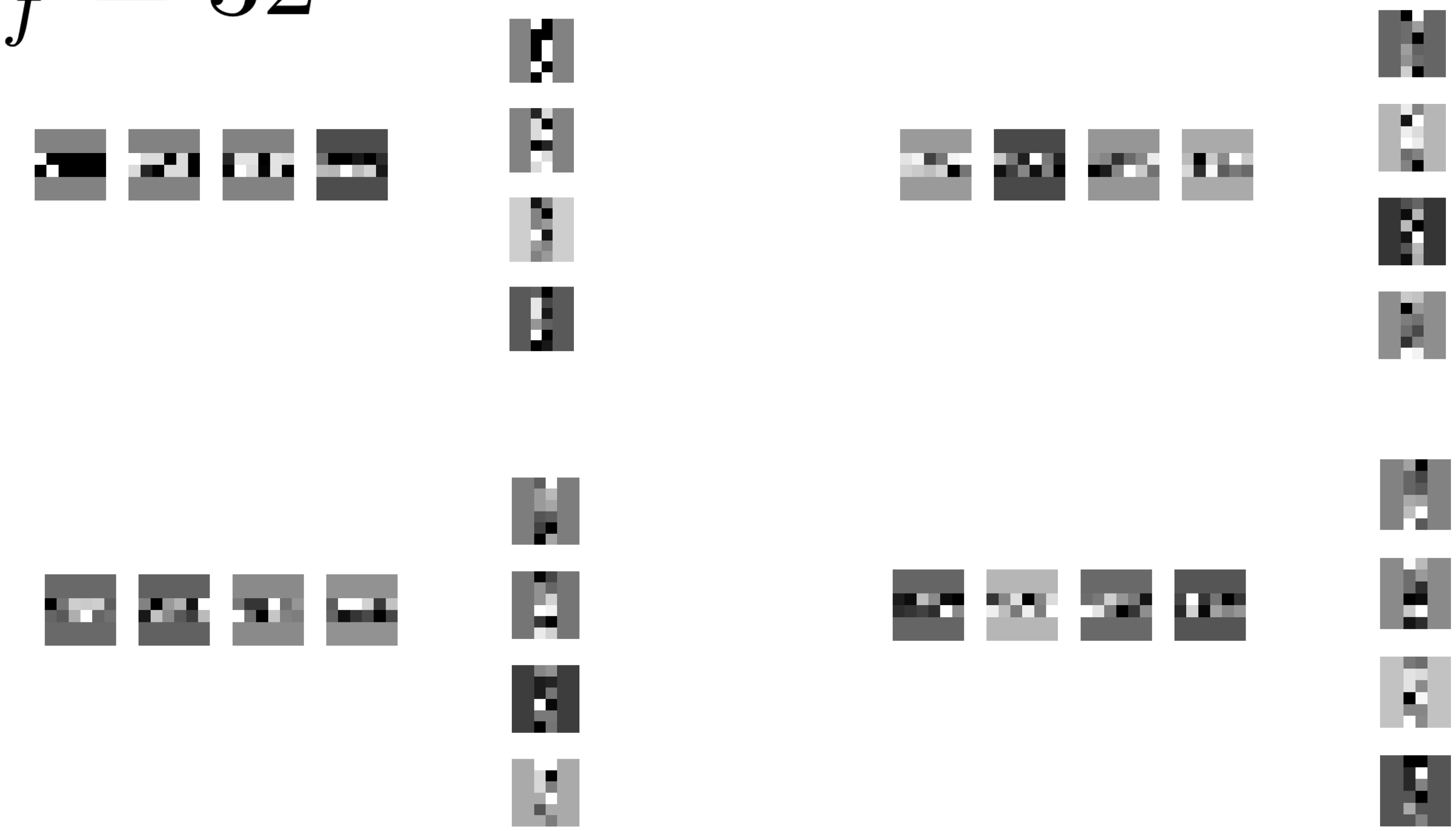
$$\boxed{8 \leq N_f \leq 1152}$$

<u>Apply *Gram-Schmidt* Algorithm:</u>

- **to fulfill the *Chimera* orthogonality**
- **the way N_f is generated defines architecture of the mapping**

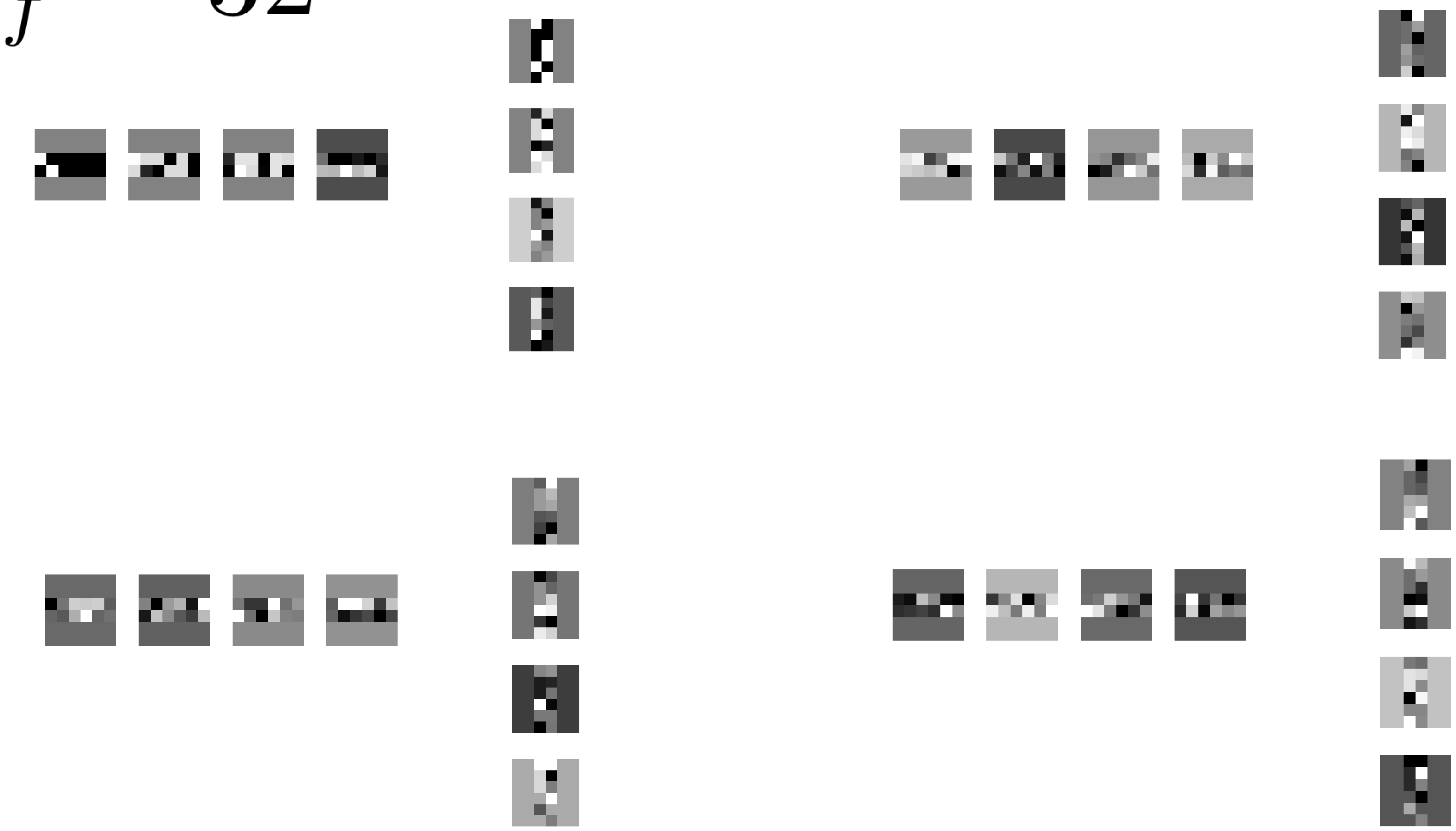
Building features

$$N_f = 32$$



Building features

$N_f = 32$

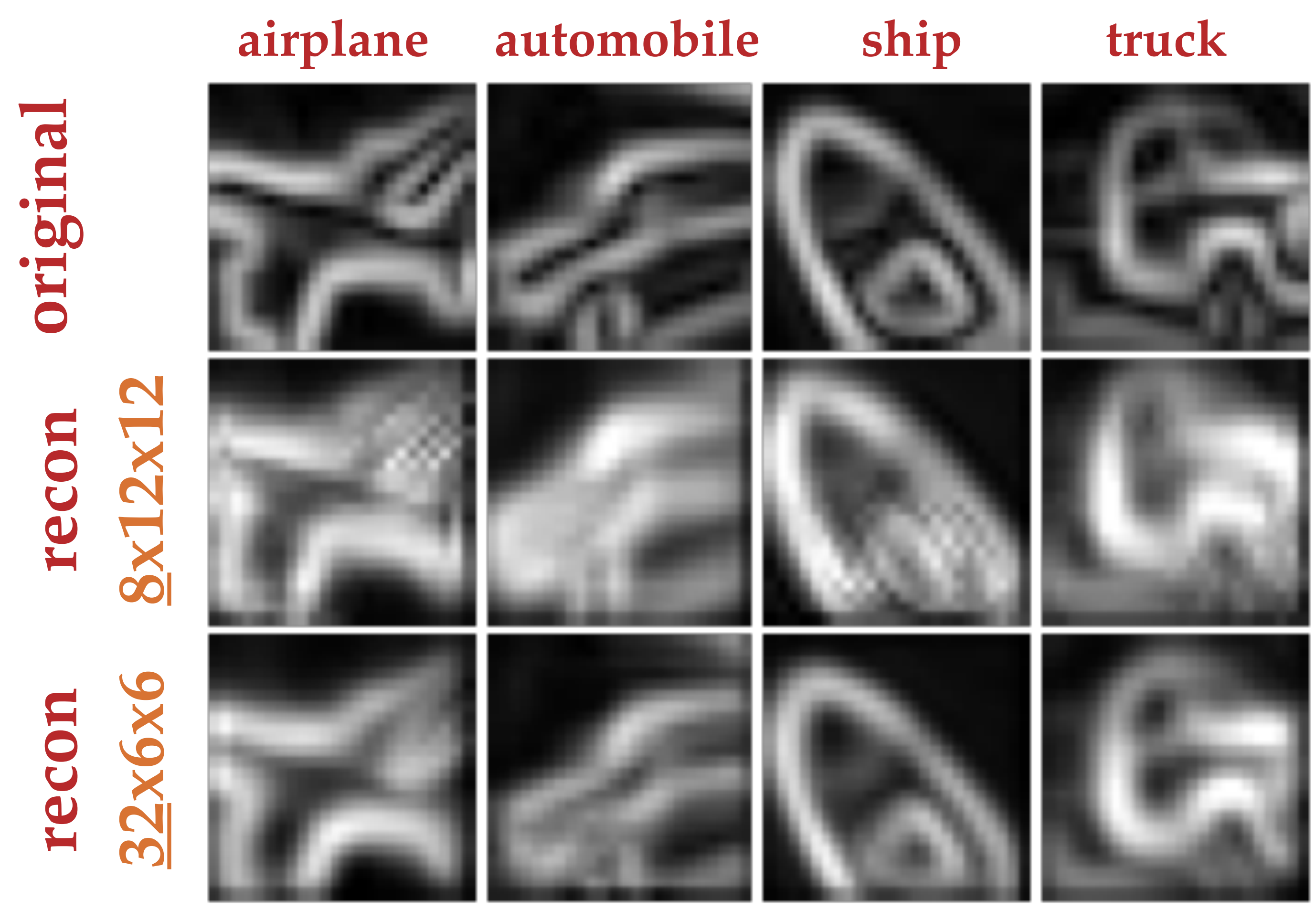


$\cdots$

$N_f = 1152$

24x24 patch images

8 and 32 features



24x24 patch images

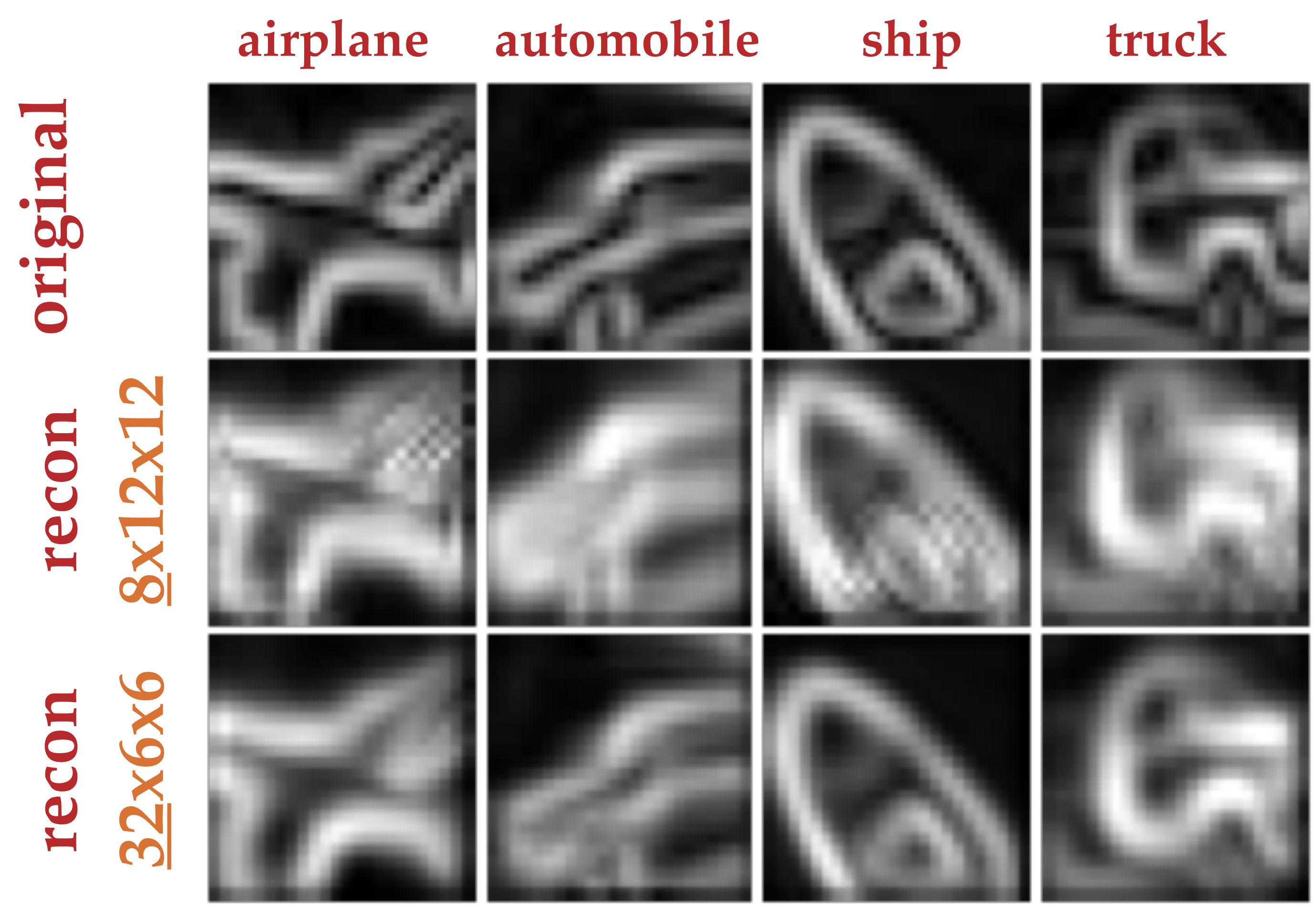
8 and 32 features

1100 *active* qubits
3068 *coupling* strengths

overcomplete order:

$$2 = \frac{12\text{x}12\text{x}8}{24\text{x}24\text{x}1}$$

stride: 2, 4



24x24 patch images

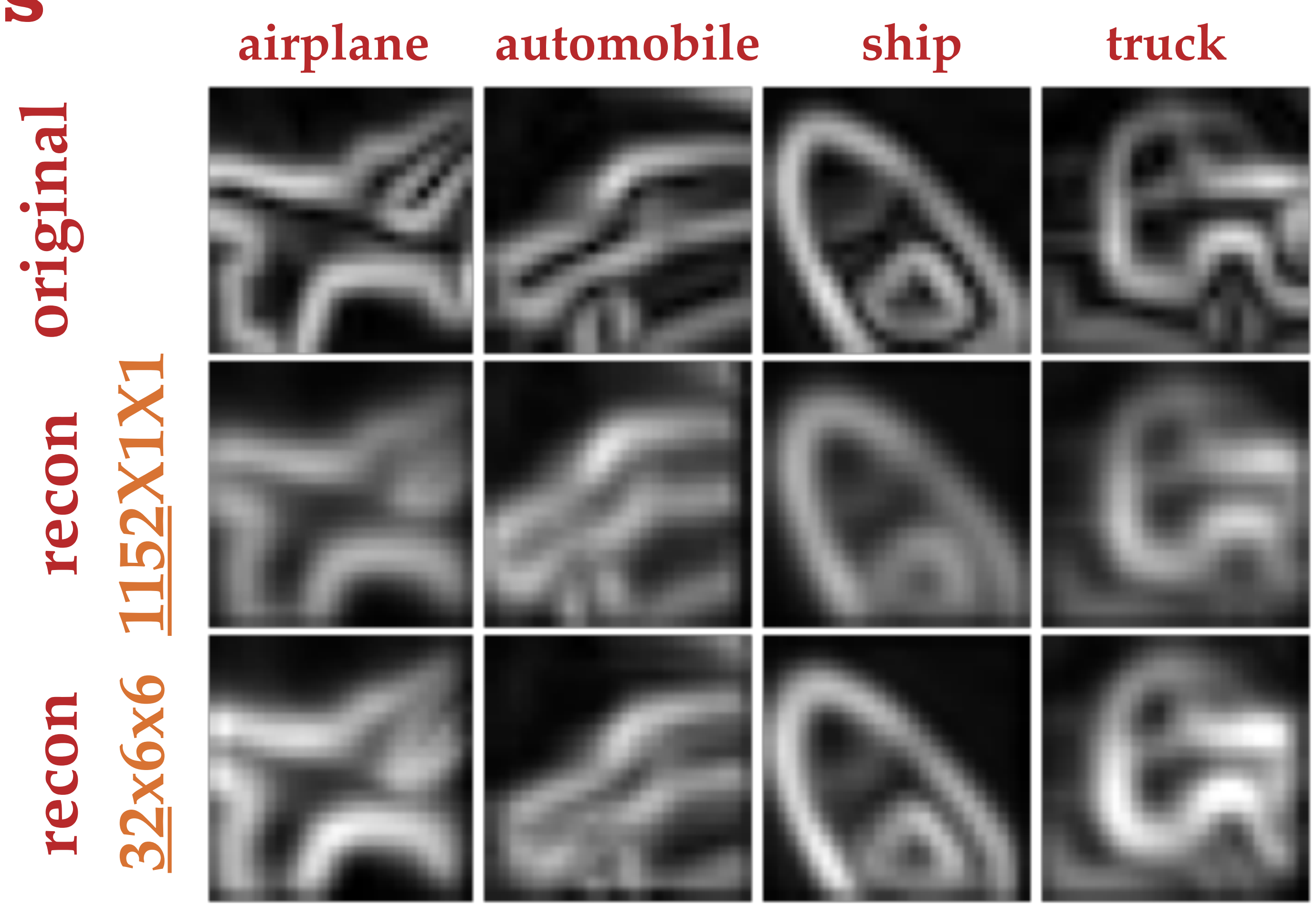
32 and 1152 features

1100 *active* qubits
3068 *coupling* strengths

overcomplete order:

$$2 = \frac{12\text{x}12\text{x}8}{24\text{x}24\text{x}1}$$

stride: 24, 4



Energy

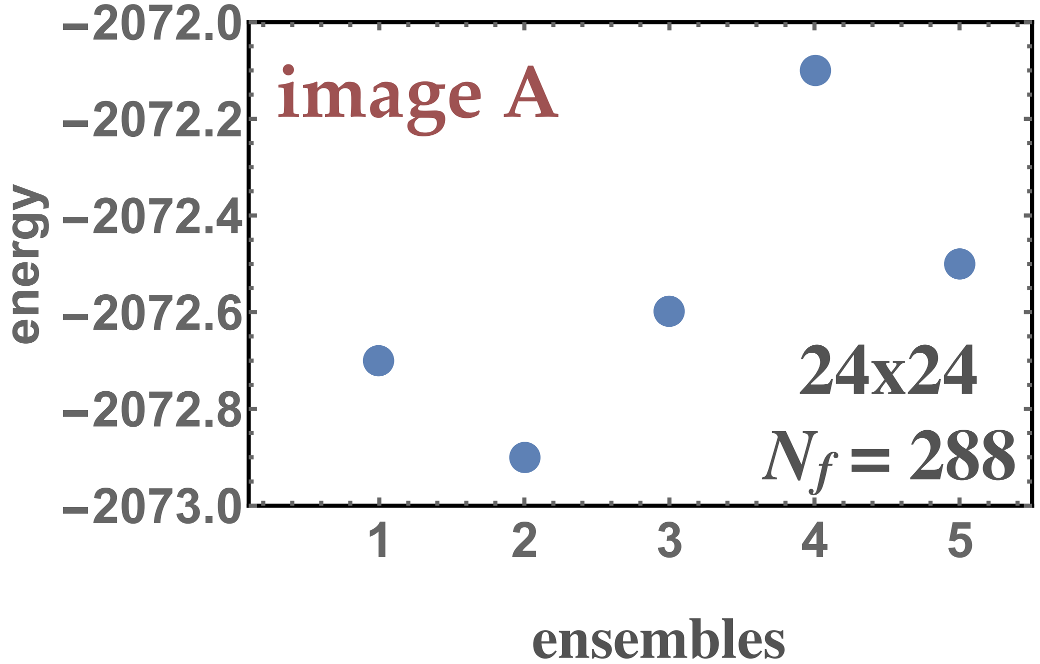


image A
24x24
$N_f = 288$
energy
ensembles
−2072.0
−2072.2
−2072.4
−2072.6
−2072.8
−2073.0
1
2
3
4
5

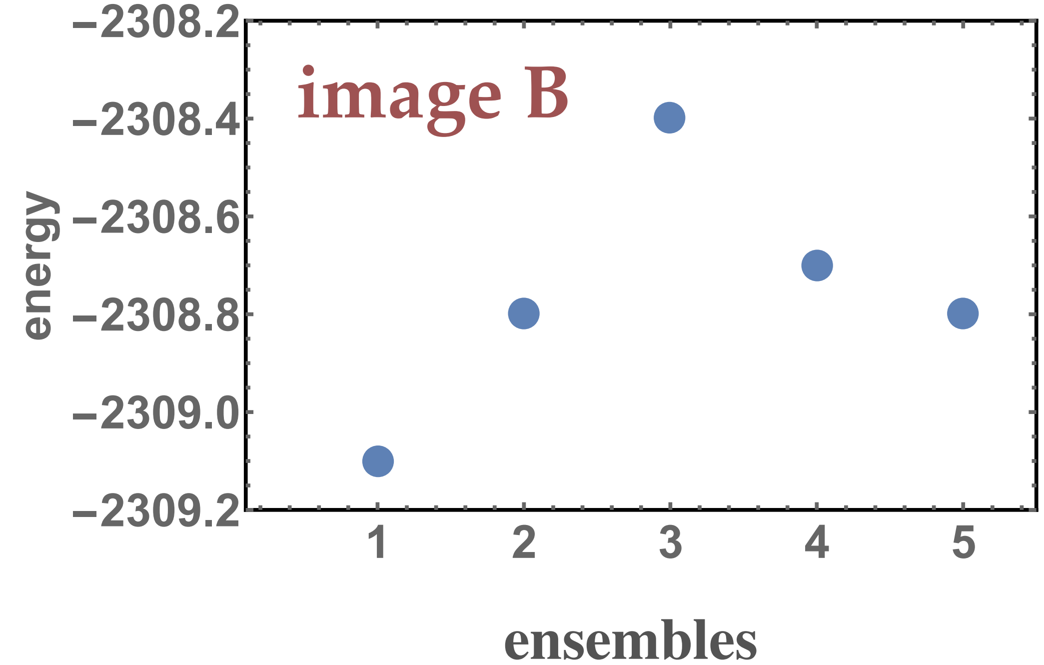


image B
energy
ensembles
−2308.2
−2308.4
−2308.6
−2308.8
−2309.0
−2309.2
1
2
3
4
5

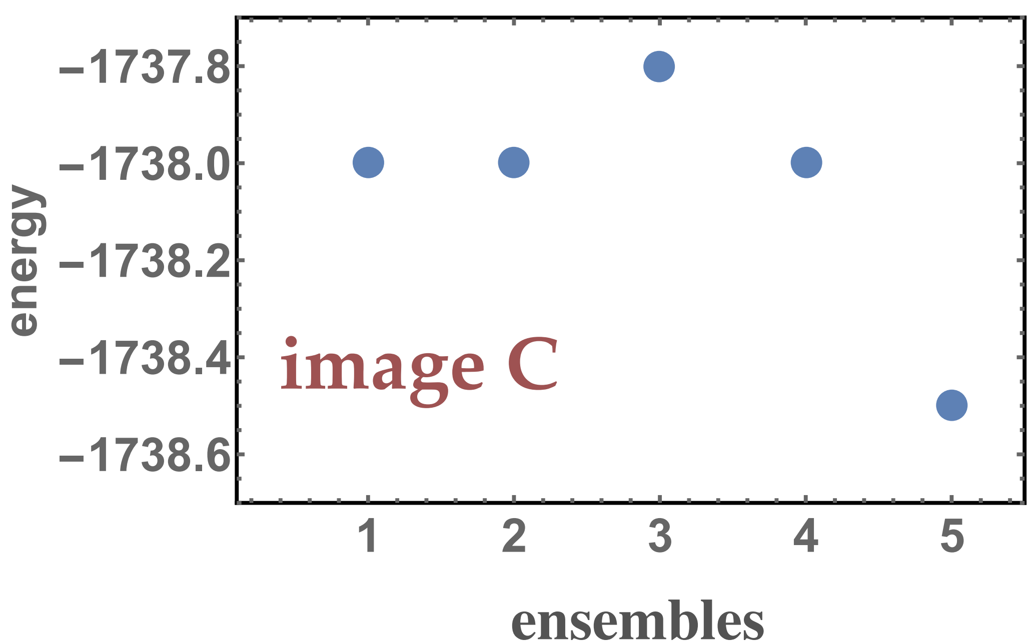
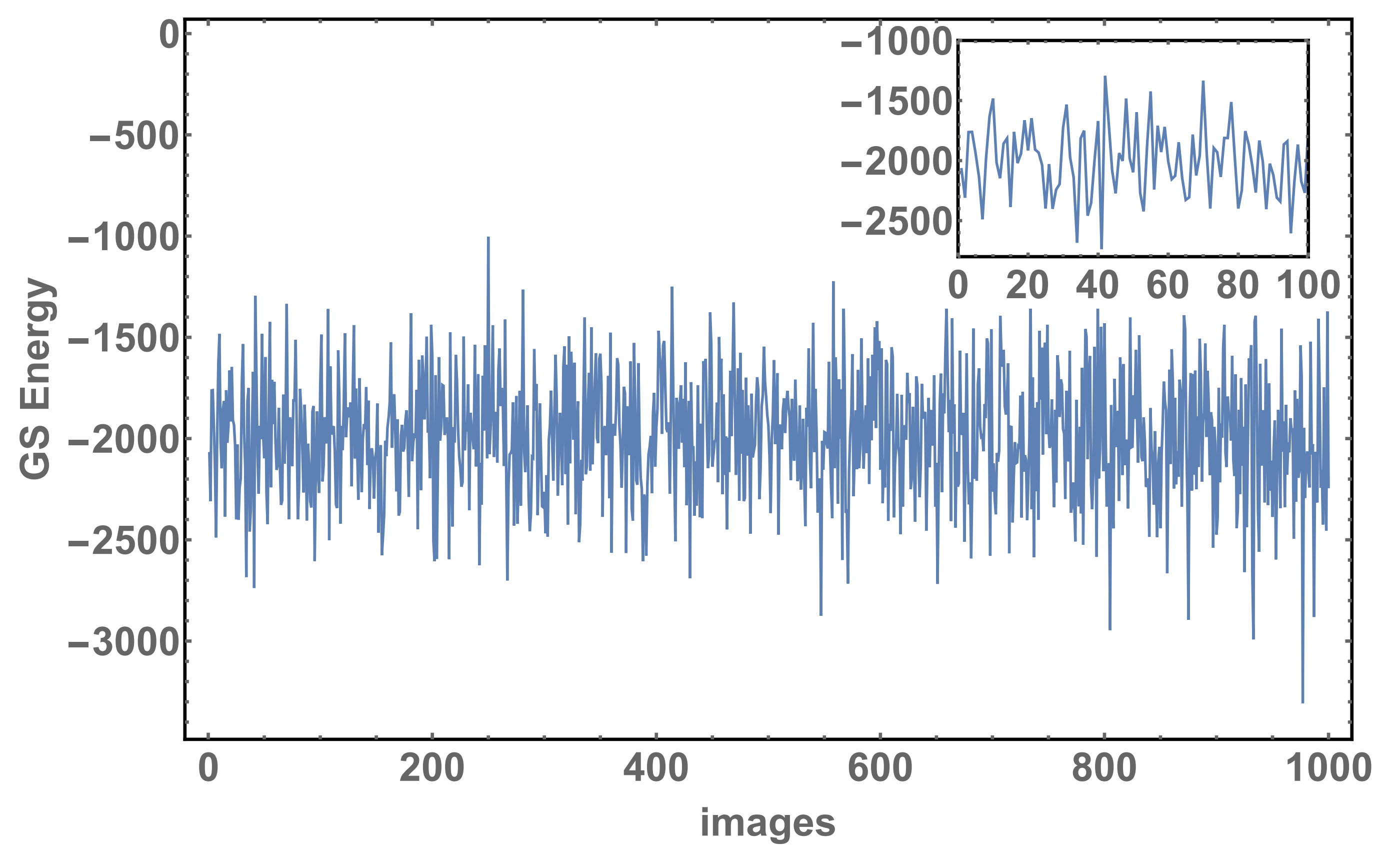
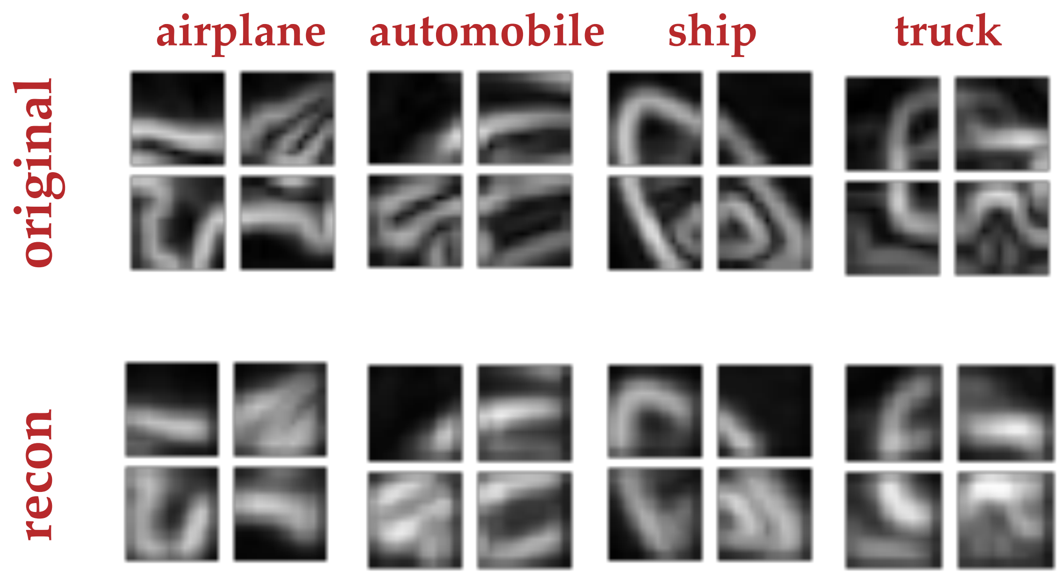


image C
energy
ensembles
−1737.8
−1738.0
−1738.2
−1738.4
−1738.6
1
2
3
4
5


GS Energy
images
0
−500
−1000
−1500
−2000
−2500
−3000
0
200
400
600
800
1000
−1000
−1500
−2000
−2500
0
20
40
60
80
100

CLASSIFICATION RESULTS

12x12 patch images



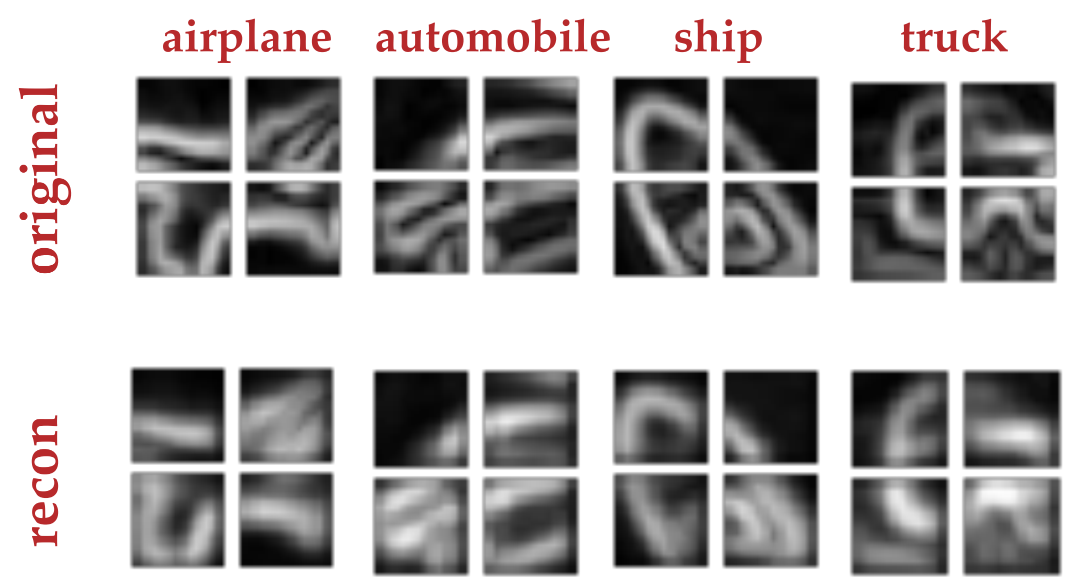
overcomplete order : 8

stride: 2

$N_f = 32$

CLASSIFICATION RESULTS

12x12 patch images



overcomplete order : 8

stride: 2

$N_f = 32$

Classification task: SVM (liblinear)

1042 training/208 test images

classes	air	auto	bird	cat	deer	dog	frog	horse	ship	truck
accur. (binary)	89.21%	93.38%	90.87%	89.42%	94.71%	88.94%	87.98%	89.9%	89.9%	85.58%

Nguyen and Kenyon, PMES-16 (2016)

COMPARISON WITH A CLASSICAL SOLVER

- So far, quantum computation (D-Wave 2X) has NOT outperformed its classical counterpart (GUROBI). Both are *comparable*.
- We already made the problem hard. We need to make it harder.
- How can we make the SC problem harder for both?

COMPARISON WITH A CLASSICAL SOLVER

From SC perspective: more overcomplete, harder to solve…

Meanwhile: The full Chimera in D-Wave offers a certain set of (*nearest-neighbor*) connectivity…

COMPARISON WITH A CLASSICAL SOLVER

GUROBI OPTIMIZATION

Los Alamos NATIONAL LABORATORY EST. 1943

From SC perspective: more overcomplete, harder to solve…

Meanwhile: The full Chimera in D-Wave offers a certain set of (*nearest-neighbor*) connectivity…

EMBEDDING technique

COMPARISON WITH A CLASSICAL SOLVER

GUROBI OPTIMIZATION

Los Alamos NATIONAL LABORATORY EST. 1943

From SC perspective: more overcomplete, harder to solve…
Meanwhile: The full Chimera in D-Wave offers a certain set of (*nearest-neighbor*) connectivity…

EMBEDDING technique

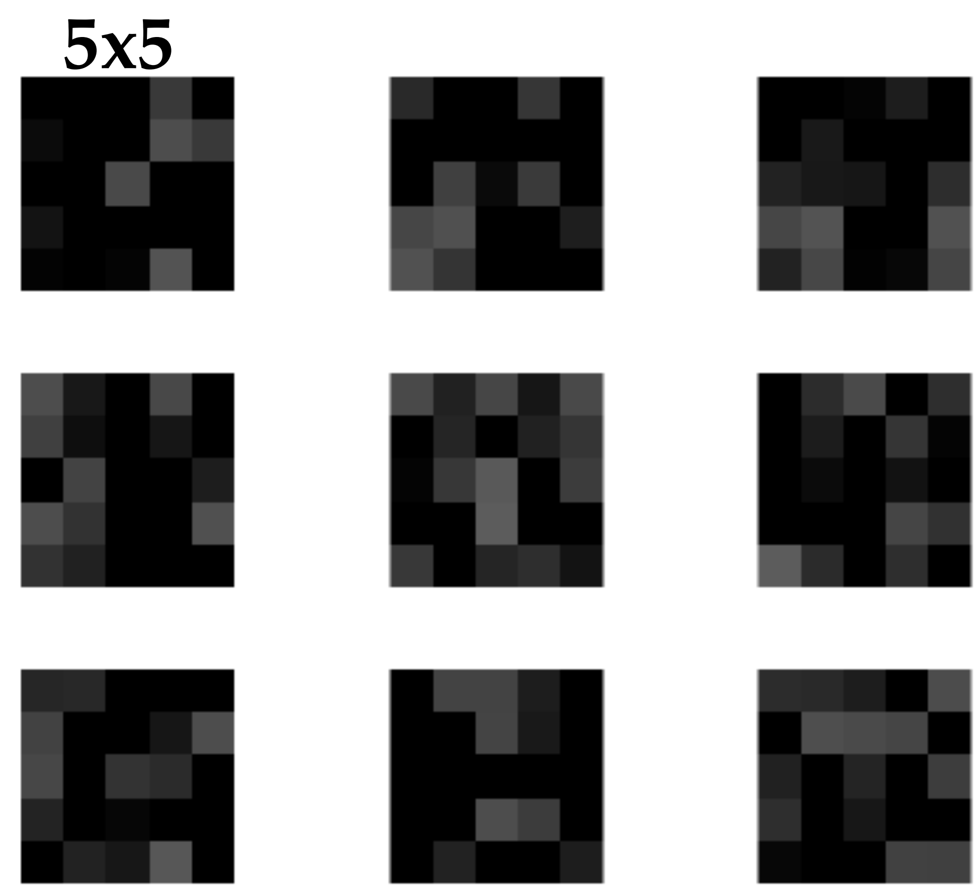
- Embedding exploits the ability to tie qubits together
- Employ all bipartite couplings
- Small number of nodes (qubits) but *more couplings* for neurons

COMPARISON WITH A CLASSICAL SOLVER

From SC perspective: more overcomplete, harder to solve…

Meanwhile: The full Chimera in D-Wave offers a certain set of (*nearest-neighbor*) connectivity…

EMBEDDING technique



COMPARISON WITH A CLASSICAL SOLVER

GUROBI OPTIMIZATION

From SC perspective: more overcomplete, harder to solve…

Meanwhile: The full Chimera in D-Wave offers a certain set of (*nearest-neighbor*) connectivity…

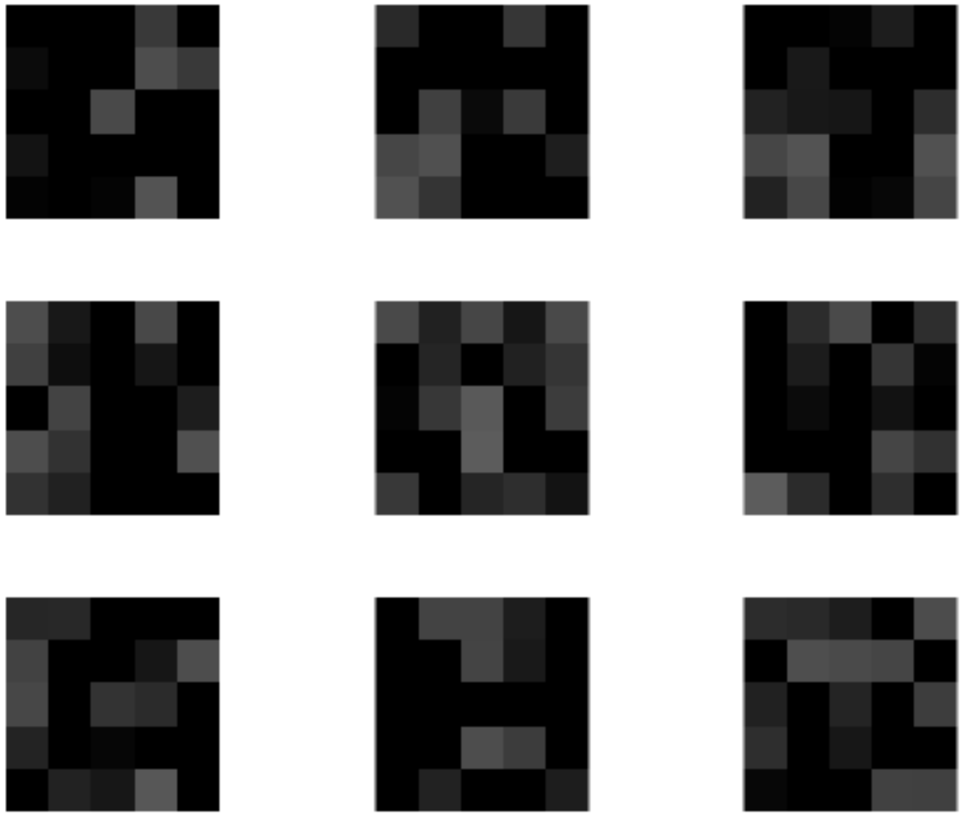
EMBEDDING technique

In practice (D-Wave 2X):

- Fully connected: 48, 49 nodes on DW2X and DW2X_VFYC, respectively
- Partially orthogonal: 72 nodes

Feature optimization!

5x5



COMPARISON WITH A CLASSICAL SOLVER

GUROBI OPTIMIZATION

Los Alamos NATIONAL LABORATORY EST. 1943

STARTING TO SEE SOMETHING GOOD...

No. of Hamiltonians: 1

solver / problem	GUROBI (best classical solver)		D-Wave 2X (ISING)	
	Energy	*Time*	*Energy*	*Time*
47 nodes: *fully* connected	-27.84	~ 300 seconds	-27.84	< 60 seconds

COMPARISON WITH A CLASSICAL SOLVER

GUROBI OPTIMIZATION

Los Alamos NATIONAL LABORATORY EST. 1943

STARTING TO SEE SOMETHING GOOD...

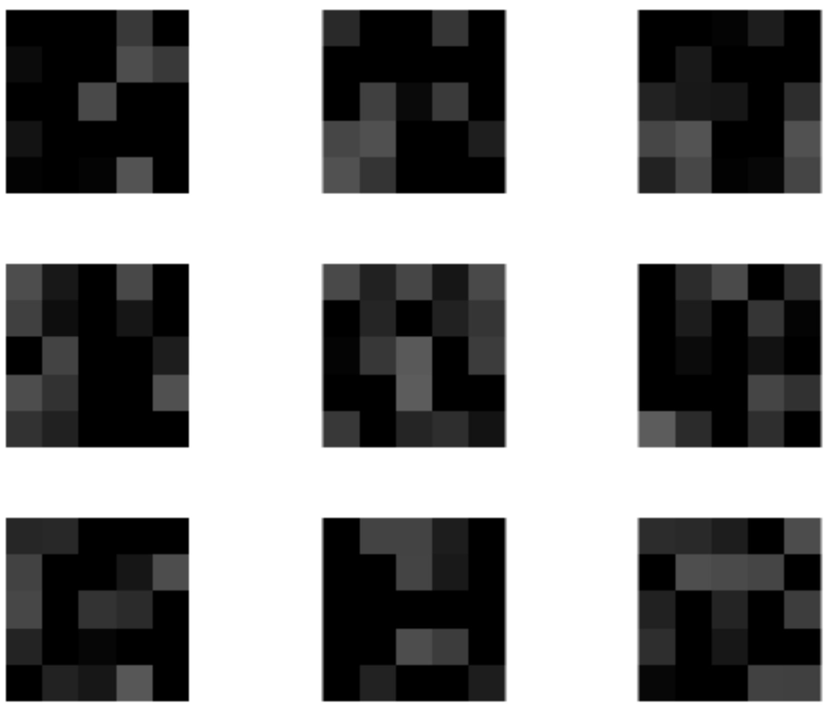
No. of Hamiltonians: 1

solver / problem	GUROBI (best classical solver)		D-Wave 2X (ISING)	
	Energy	Time	Energy	Time
47 nodes: *fully* connected	-27.84	~ 300 seconds	-27.84	< 60 seconds
	Energy	Time	Energy	Time
70 nodes: *partially* Chimera-orthogonal	-43.251	~2000 seconds	-43.251	< 60 seconds

Feature Learning (in progress)

before…

5x5



feature optimization

Stochastic gradient descent

{ given a set of neuron activity $\vec{a}$ generated by D-Wave 2X,

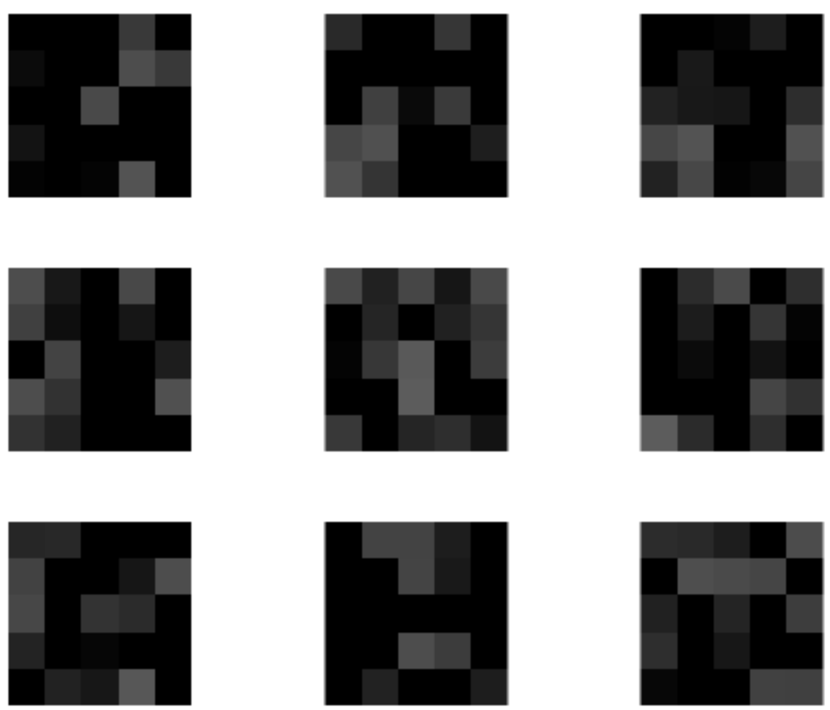
```
    do:

        for iteration
            for mini_batch %[1:size(sampling)]
                %update weights
                φ := φ − η ▽ E(φ)
            end
        end

    end
}
```

$$\phi := \phi - \eta \bigtriangledown E(\phi)$$

Feature Learning (in progress)

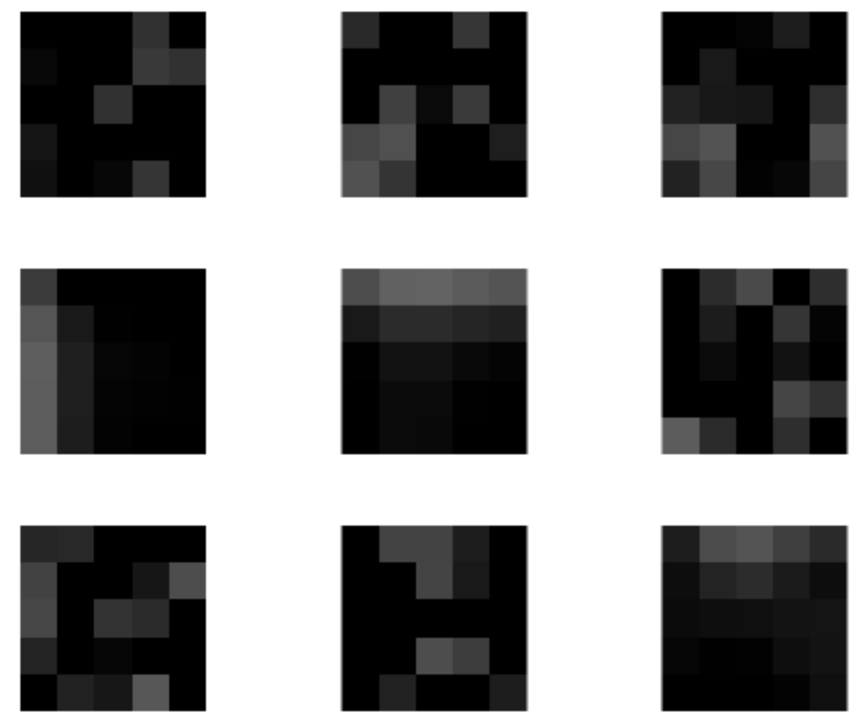
before…

5x5



…after

5x5

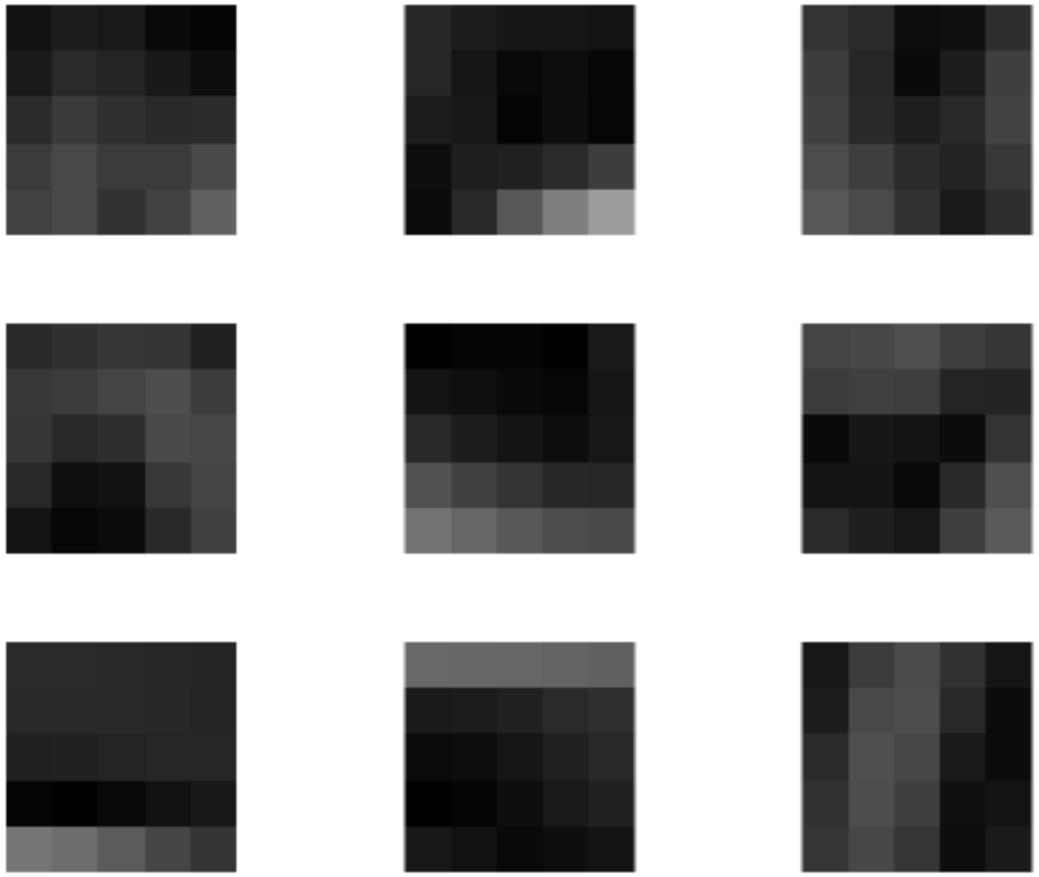


many "*lazy*" features

...THE UNEXPECTED
Imprinting technique

GENERATING FEATURES

randomly sampled *imprinting* features

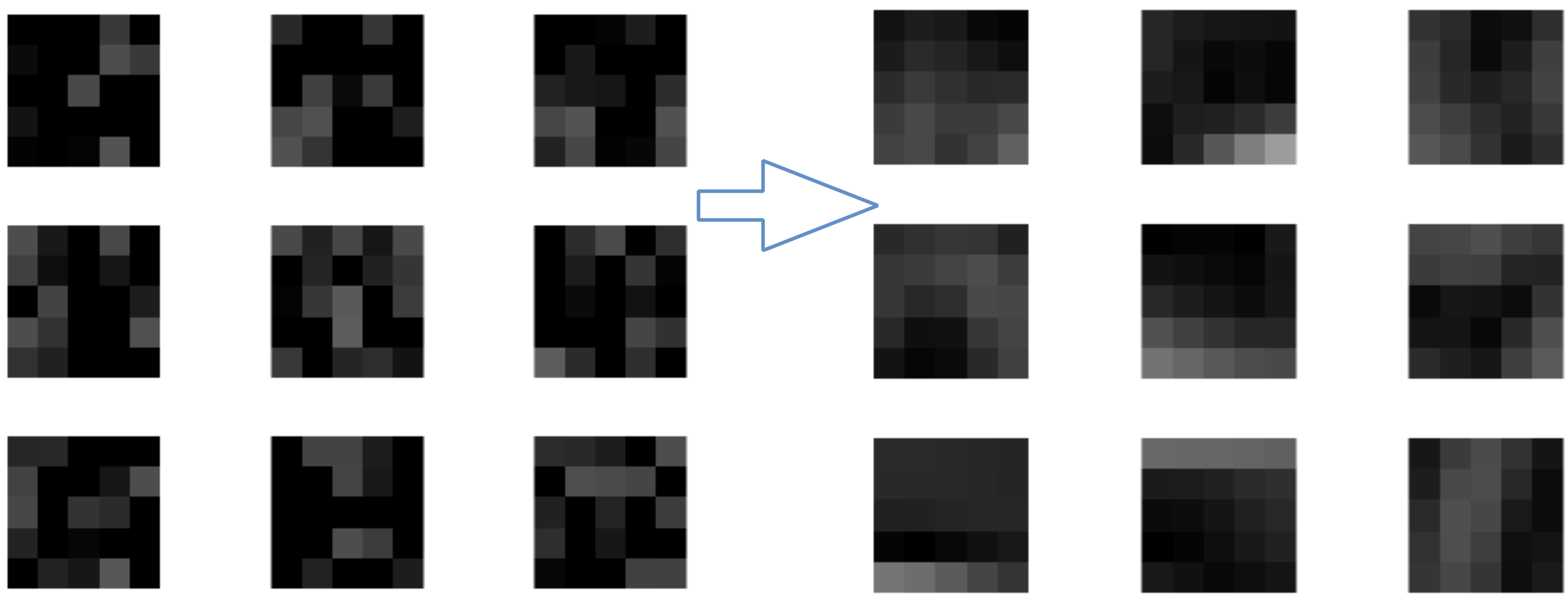


...THE UNEXPECTED

Imprinting technique

GENERATING FEATURES

randomly generated features | randomly sampled *imprinting* features



Does this enhance the "*hardness*"?

...THE **GREAT!** UNEXPECTED

Imprinting technique

solver / problem	GUROBI (best classical solver)		D-Wave 2X (ISING)	
Sparse coding	*Energy*	*Time*	*Energy*	*Time*
	-129.533	(*cutoff*) ~ 9 hours	*-131.14*	< *60 seconds*

...THE GREAT! UNEXPECTED

Imprinting technique

Feature learning

...THE GREAT! UNEXPECTED

Imprinting technique

Feature learning

100% adaptive features

OUTLINE

A. SPARSE CODING ON A QUANTUM D-WAVE

B. IMPLEMENTATION ON D-WAVE MACHINE

C. SPARSE CODING FOR OBJECT DETECTION

D. **SUMMARY AND FUTURE WORK**

D. SUMMARY

- **first demonstration of *sparse coding* using a quantum computer**
- **mapping of visual features to D-Wave 2X Chimera**
- **benchmark results on standard image classification task**
- ***compare* D-Wave 2X performance with GUROBI**
- **obtained solutions to the problems where D-Wave 2X *significantly* outperforms GUROBI**

work in progress…

CIFAR-10

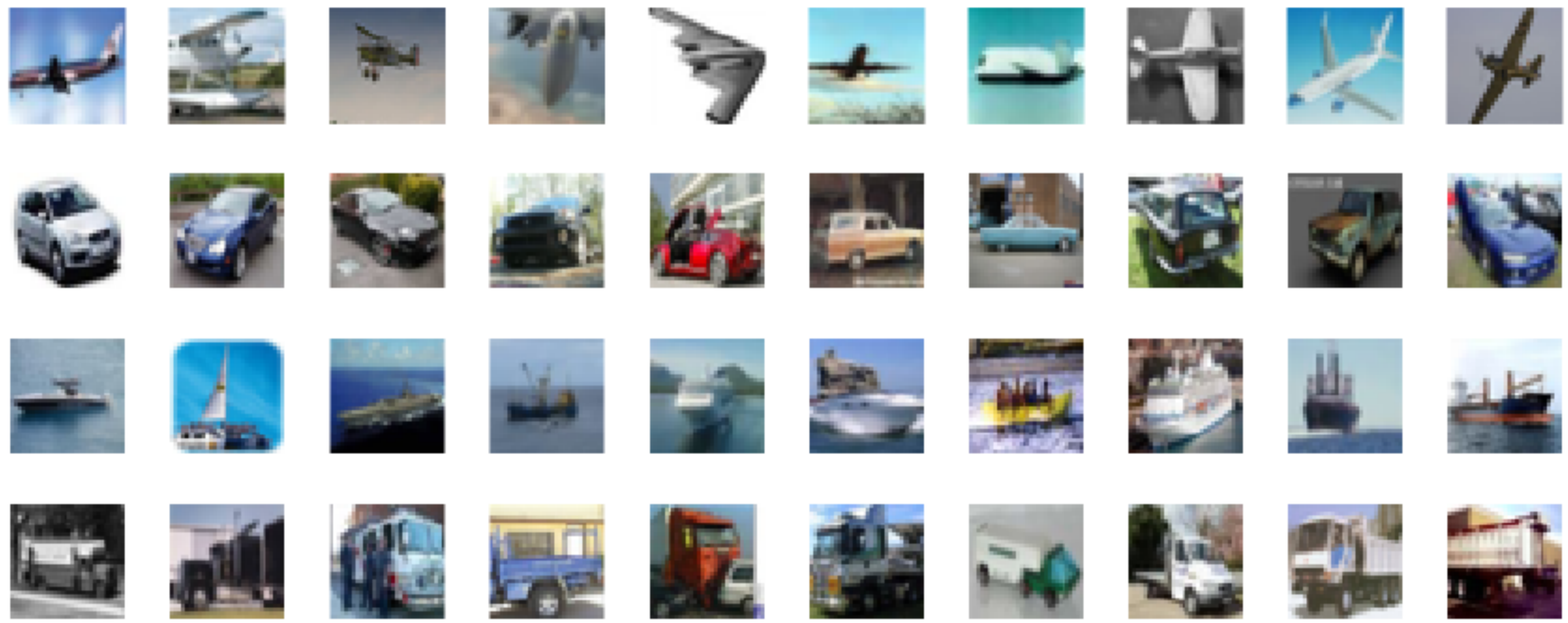
32x32

airplane

automobile

ship

truck



30x30

edge

color

D. (IN PROGRESS &) FUTURE WORK

- optimize features
- add colors
- hierarchy model
- TrueNorth comparison